

CHAPTER 56

SEA STATE PARAMETERISATION USING EMPIRICAL ORTHOGONAL FUNCTIONS

Witold Cieřlikiewicz¹ and Jerzy Graff¹

Abstract

In this paper the development of a parametric transformation linking the meteorological conditions with the sea response for the Irish Sea and the southern North Sea regions is presented. The method of *empirical orthogonal functions* and *system identification* procedures are used to develop the parameterisation. The reference data sets used in the investigation consist of high quality hindcast data (41 storm events) and field measurements (13 years) covering meteorological and sea state parameters. The method of empirical orthogonal functions is applied to the wind velocity field and it is shown that the wind field time history can be adequately represented by the first few principal components. These wind field principal components together with atmospheric pressure parameters are used to synthesise the meteorological input for system identification. The system identification procedures are then applied to develop a new efficient form of parametric model linking spatial meteorological data with sea state response.

1. Introduction

This study is a part of the NEPTUNE project (see Graff & Cieřlikiewicz 1996, and Graff *et al.* 1995) under the EU MASTII framework. It is concerned with the requirement to develop a fast and efficient scheme to predict the sea state response during extreme storms using wind field and atmospheric pressure field data. The project methodology involves a cause-effect process chain, the first part of which links the storm meteorological variables field $\mathbf{M}_t$ and the offshore sea state variables $\mathbf{X}_t$. In the project a relatively long data set is available for meteorological variables $\mathbf{M}_t$. This is the *Norwegian Meteorological Institute—DNMI* 6-hourly gridded pressure field $\mathbf{p}_t$ and wind field $\mathbf{W}_t$ for the North Sea – NE Atlantic covering the period 1955–1993. In order to create long time series of offshore variables $\mathbf{X}_t$ the parametric transformation $\mathcal{F} : \mathbf{M}_t \rightarrow \mathbf{X}_t$ has to be developed. In this paper we present a new approach to derive the parametric transformation models $\mathcal{F}$.

There are many classic methods for prediction of sea state determined by local winds under prescribed conditions of fetch and wind duration. The simplest assume a uniform and steady wind blowing over a limited fetch or over an unlimited ocean for time t after a sudden onset. These are cases of fetch-limited and duration-limited waves, respectively. Those methods provide a very useful first look at the wave field,

¹ British Maritime Technology Limited, 7 Ocean Way, Ocean Village, Southampton SO14 3TJ, England

but the assumption of a uniform wind field is generally unrealistic and is often a source of considerable error. Examination of regional wind fields over relatively small areas, e.g. the Irish Sea, during hurricane type storms shows that the wind field is not uniform. The variation of both the magnitude and the direction of wind velocity vector, which influences the wave field, is small but significant. The spatial pattern of the wind velocity field reflects the circulation of the air in the cyclonic direction around the centre of a moving pressure depression (see Fig. 1). Over the larger areas, the non-uniformity of the wind field is much more pronounced and the sea state wave forecasts based on fetch-limited or duration-limited algorithms cannot be usually accepted. A further limitation of these methods is their difficulty in reflecting the *dynamic* features of the metocean system, i.e., the fact that the current sea state depends not only on the current meteorological conditions but also on their history.

In this study we seek a fast and efficient prediction scheme as an alternative to conventional numerical modelling. The scheme must be capable of taking into account the non-uniformity of the wind field and should incorporate the most characteristic features of its spatial distribution concerning storm track history. Moreover it should also reflect the dynamic features of the metocean system.

When long historical time series of metocean parameters are available it becomes possible to apply *system identification* (SI) techniques to investigate the required types of prediction schemes. In this study, the available data consist of the DNMI 6-hourly gridded meteorological data $\mathbf{W}_t$, $\mathbf{p}_t$ and the offshore variables $\mathbf{X}_t$ describing the sea state in terms of both hindcast data (41 storm event periods) and 3-hourly field measurements extending over 13 years 1979–1991 at five monitoring stations near the Dutch coast (provided by the *National Institute for Coastal and Marine Management of Rijkswaterstaat—RIKZ*). The parameters selected from the RIKZ data set consist of time series of significant wave height $H_S(t)$, mean wave period $T_z(t)$, principal wave direction $\theta_0(t)$, astronomical tide and the observed still water level. The difference between the latter two was taken as the surge level $S(t)$.

The application of the SI technique in sea response modelling, when the non-uniform wind field over selected area is taken as the input data, will lead to very large model dimensions and to enormous amounts of data involved. For example, in this study for the southern part of the North Sea the area covered by a sector of 221 DNMI grid points was selected. This leads to 442 time series of wind components, each 37984 data points long. This amount of input information which has to be processed in the

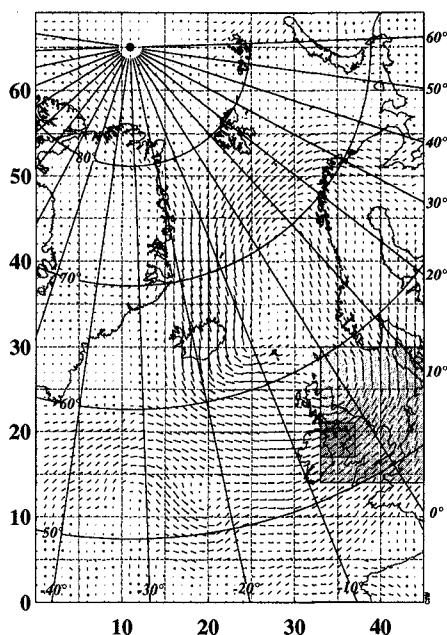


Fig. 1. DNMI wind field at elevation 10 m at 0:00h on 12 Nov 1977. DNMI grid coordinates and study areas.

SI modelling may be well beyond the capabilities of most desktop computer systems available nowadays. However, the question arise: is all that input information equally important? Certainly not. It will contain noise and, moreover, it is possible that part of that information is in a some way redundant. In order to reduce the volume of input wind data and still preserve its salient features, while filtering out most of the noise and redundant information, the *empirical orthogonal functions* (EOF) analysis is used. EOF analysis in time domain, has been variously used in meteorological, environmental, and oceanographic studies for some two decades (see e.g. Prandle & Matthews 1990, Ng 1993). It is applied here to analyse and decompose the DNMI wind field time series over two selected areas, namely the Irish Sea and the southern North Sea.

In the EOF analysis, the temporal and spatial variation of the wind field is partitioned into orthogonal spatial patterns, so called EOF modes $\mathbf{e}_m$, which are constant in time domain and principal components $P_m(t)$ (corresponding to each EOF mode). Each principal component $P_m(t)$ is given as a time series describing the time evolution of the corresponding EOF mode. The importance to this study is the fact that the wind field time history can be determined with sufficient accuracy by a few principal components only. Those few principal components and not the wind time series itself, are taken as the SI stimuli and are used to build the parametric models linking the meteorological parameters field $\mathbf{M}_t$ with the offshore sea state variables $\mathbf{X}_t$.

We assume that input-output variables are related by a *linear dynamic system* and we use SI models in multi-input, single-output version, i.e. each component of the offshore parameters vector $\mathbf{X}_t$ is modelled one by one. The estimated models $\mathcal{F}_{\mathcal{L}}^X : \mathbf{Q}_t^X \rightarrow X_{\mathcal{L}}$ link the meteorological input $\mathbf{Q}_t^X$ with the output $X_{\mathcal{L}}$ which is one of the sea state parameters: $H_S(t)$, $T_z(t)$, $\theta_0(t)$ or $S(t)$ at offshore monitoring station indicated by the location index $\mathcal{L}$. The input $\mathbf{Q}_t^X$ is composed using the first few wind field components $P_m(t)$ and the atmospheric pressures p_i in the selected grid points indexed by i . We use superscript X to emphasise the fact that the model input may be prepared differently for different output variables. For example, when the surge is modelled we use atmospheric pressure, which is however not the case for the wind wave variables.

Summarising, there are two important elements in this study: The EOF analysis, and SI procedure. Both are schematically depicted on Fig. 2. The EOF prepares the input data by reducing the amount of input information, i.e., by extracting only that information which is most significant. Then, the SI deals with the actual problem of building a mathematical model of a dynamical metocean system in which the sea response is stimulated by the meteorological conditions.

2. EOF analysis of wind field

2.1 Basics

Let $\mathbf{W}_t$ denote the complex state vector formed by the M complex functions of time $W_m(t)$

$$W_m(t) = U_m(t) - \langle U_m \rangle + i(V_m(t) - \langle V_m \rangle) \quad (1)$$

where $U_m(t)$, $V_m(t)$ are the wind velocity components in m th of M locations and $\langle \cdot \rangle$ denotes the expected value of a quantity enclosed in the angle brackets. As the covariance matrix $\mathbf{H}$

$$\mathbf{H} = \langle \mathbf{W}_t \mathbf{W}_t^+ \rangle \quad (2)$$

is Hermitian (the cross denotes the transpose complex conjugate) it has M real eigenvalues λ_m and complex unitary eigenvectors $\mathbf{e}_m$ which are called *EOF modes* and may be normalised: $\mathbf{e}_m^+ \mathbf{e}_n = \delta_{mn}$

The EOF modes $\mathbf{e}_m$, as eigenvectors, form a complete and orthonormal basis for $\mathbf{W}_t$. Thus the original wind field state vector $\mathbf{W}_t$ may then be expanded in terms of the EOF modes

$$\mathbf{W}_t = \sum_n P_n(t) \mathbf{e}_n \quad (3)$$

The so called *principal components* $P_m(t)$ are obtained as

$$P_n(t) = \mathbf{e}_n^+ \mathbf{W}_t \quad (4)$$

It can be shown that the principal components compose a set of orthogonal vectors satisfying the relation $\sum_{\nu=1}^N P_{\nu}^* P_{\nu} = N \lambda_n \delta_{nm}$ which shows the principal components corresponding to different

modes are uncorrelated in time and that $\lambda_m = (\sum_{\nu=1}^N |P_{\nu m}|^2)/N$ is the mean energy contained in the EOF mode $\mathbf{e}_m$. The fraction χ_m of the total variance (total in the sense that this is the local variance of wind velocity summed over all locations, i.e. $\text{Tr}(\mathbf{H})$) corresponding to the EOF mode $\mathbf{e}_m$, with the eigenvalue λ_m , is given by

$$\chi_m = \lambda_m / \text{Tr}(\mathbf{H}) \quad (5)$$

We shall assume henceforth that EOF modes are ranked in descending order according to that fraction.

Concluding, the EOF analysis separates the space-time variation of the wind field state vector into the space variation of the EOF modes which are constant in time and uncorrelated over space and the time variation of the principal components that do not depend on location in space and are uncorrelated in time. The time evolution of each EOF mode $\mathbf{e}_m$ is described by a time series of the principal component $P_m(t)$ defined in (4). The observed wind field pattern at a given time in the study region is given by the sum of the mean wind velocity vector and EOF modes, each being modulated by the complex value of the corresponding principal component at that time. It is assumed in this study that those EOF modes that account for small fractions of total variance are not important for the physical process which is modelled. In that sense the EOF technique provides an effective way to reduce, or compress, the data and filter out most of the noise whilst retaining the most relevant information to be incorporated into the analysis.

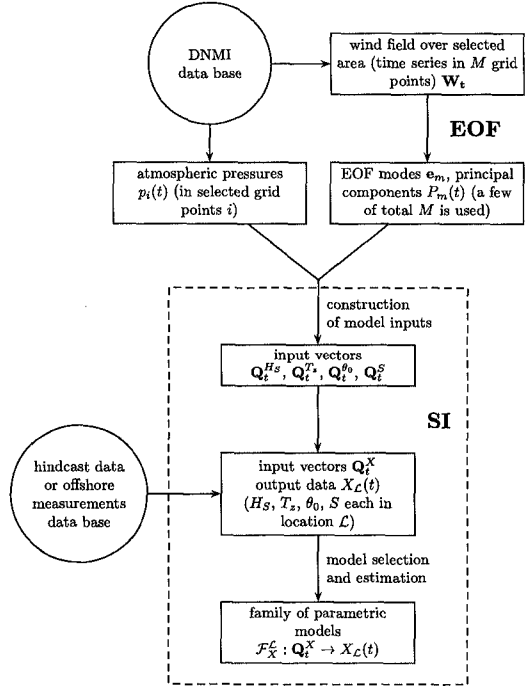


Fig. 2. Methodology for the study.

2.2 Application of EOF analysis

The DNMI 6-hourly gridded pressure field and wind field data which cover the North Sea and NE Atlantic are given in a rectangular grid on a polar stereographic projection with grid size 75 km at 60° N. The area covered by the DNMI data base and the grid used are presented on Fig. 1. In the project NEPTUNE two distinct demonstrator zones were established, one on the west coast of Great Britain and the other on the Dutch coast. The EOF analysis of wind field data was applied for both demonstrator zones. In both cases the rectangular areas for EOF analysis were selected. They are marked on Fig. 1. The area covering the Irish Sea consists of 5×5 grid points, i.e., $M = 25$. In the case of the Dutch zone, because of the open character of North Sea, much more wind time series have been subjected to the EOF analysis. The area selected for the Dutch zone consists 13×17 grid points resulting in $M = 221$.

In this section, as an example, the result of EOF analysis performed for two storm events and for two whole years over the Irish Sea is presented. Over the southern North Sea the results of EOF analysis performed for the 13 year period are described in greater detail. In Table 1 the fractions χ_m of the total variance corresponding to the first four EOF modes for the Irish Sea area are presented.

Table 1. First four EOF modes of the wind velocity field over the Irish Sea.

Mode	Percent of total variance			
	1977		1983	
	7 Nov – 17 Nov	whole year	25 Jan – 6 Feb	whole year
1	93.88	90.08	93.07	92.71
2	2.69	4.10	4.32	3.55
3	2.26	3.37	2.07	3.10
4	0.42	0.58	0.23	0.28

On Fig. 3(a) the mean wind velocity field during the storm period 7–17 November, 1977 is shown. Fig. 3(b) presents the associated first two modes. The most important first EOF mode containing 94% of total variance appeared to be almost uniform with characteristic “cyclonic twist” clearly visible. On Fig. 3(c) the time histories of the first two principal component vectors during the storm are shown. The orientation of vectors is related to DNMI grid co-ordinates.

It was found that the EOF analysis performed for each of the whole year periods results in spatial patterns very similar to those obtained for the storm event only. Of course, the mean wind field over the whole year differs significantly from that calculated for storms only. The first EOF mode still consists about 90% of the total variance (see Table 1), which suggests possible usage of the EOF technique for describing the continuous time history of the wind field over the longer period covered by DNMI data set. In the case of the southern North Sea the first four EOF modes for 6 year and 13 year periods were calculated and compared. They proved very similar and it was decided to utilise the result of EOF analysis for 13 year period in further SI work.

The period 1979–1991 covered by RIKZ offshore parameters data base was selected. To adjust the 6-hourly wind velocity time series to the 3-hourly RIKZ measurements the interpolation suitable for SI analysis was performed first. Because of the greater number of the wind time series taken into the analysis ($M = 221$) and the wider area

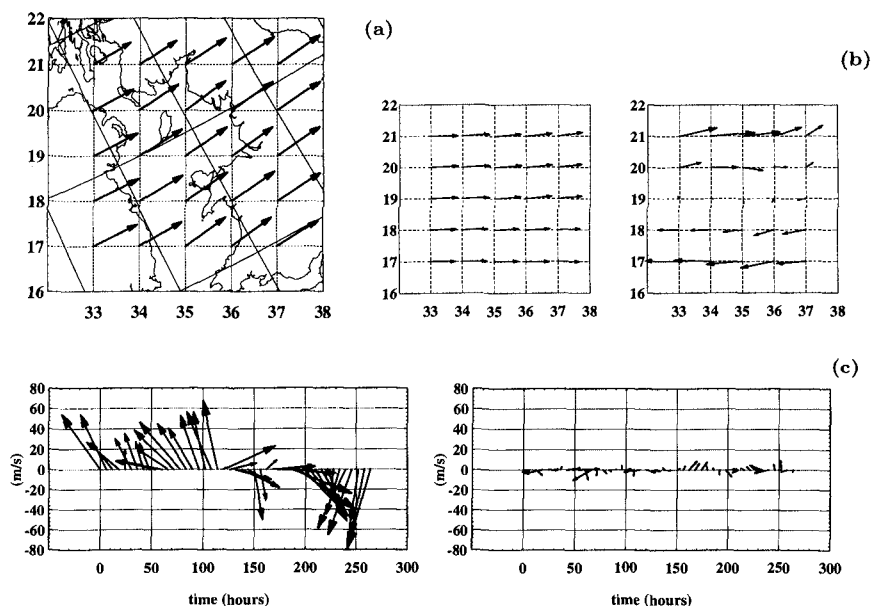


Fig. 3. EOF analysis of wind velocity field over the Irish Sea during the storm period 7 Nov – 17 Nov, 1977; (a) mean wind velocity field, scale: upper-left vector has length 10.70 m/s; (b) first (left) and second (right) EOF modes, scale: upper-left vector in the first mode has the length 0.20 (vectors of EOF modes are dimensionless); (c) evolution of first (left) and second (right) principal components, scale: left-hand side vector in the first component has the length 66.29 m/s. Time is counted starting from 0:00h of first day of storm period.

covered, the 90% of the total variance is distributed over the first eight EOF modes. The numerical values for the first eight EOF modes are given in Table 2. On Fig. 4(a) the mean wind velocity field over the area covering the 221 DNMI grid points used for the EOF analysis over 13 years is presented. Fig. 4(b) shows the first four EOF modes. Similar to the Irish Sea area case, the first mode shows the cyclonic twist characteristic for storm events determined by the depressions situated North of the Dutch coast. Also the second EOF mode is similar to the second mode of the Irish Sea case. It becomes clear that the EOF modes reflect certain characteristic features of the storm climate of both the Irish Sea and the southern North Sea areas. This is demonstrated for short storm periods and for longer one year periods as well as for the relatively long 13 year period.

Table 2. First eight EOF modes of the wind field over the southern North Sea.

Mode	1	2	3	4	5	6	7	8
Percent of total variance	59.50	15.15	9.68	3.55	2.22	1.57	1.34	1.01

On Fig. 4(c) the time series of the first four EOF principal components for one of the extreme storms selected within the study for Dutch demonstrator zone are presented for illustration.

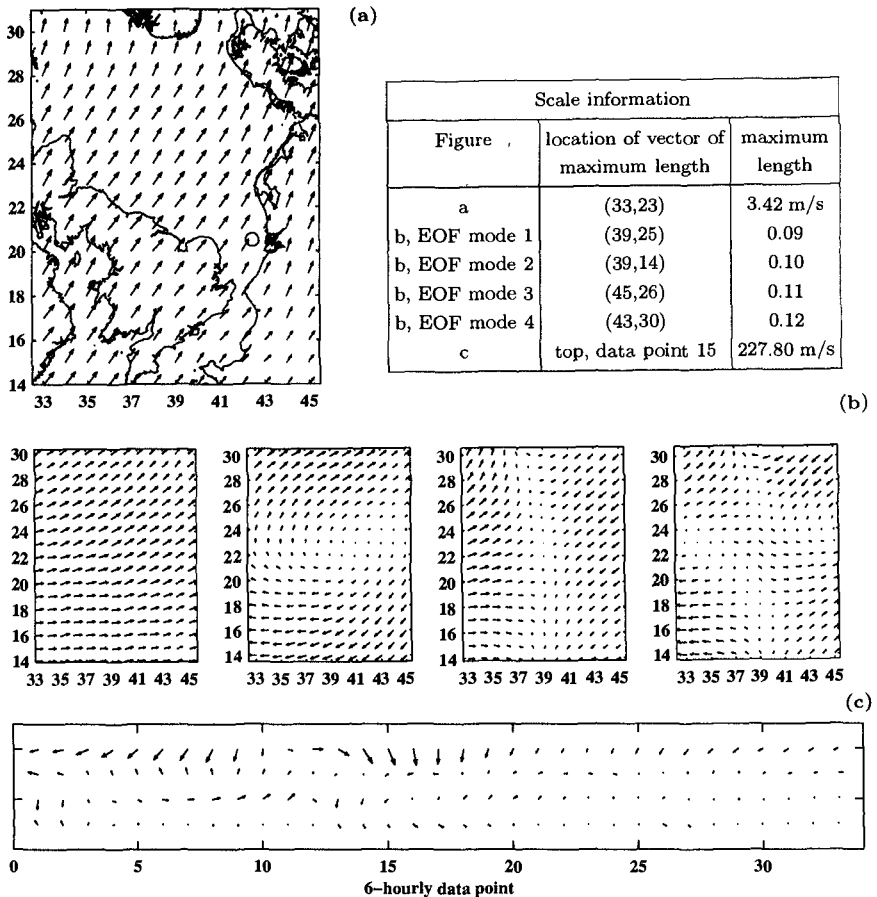


Fig. 4. EOF analysis of wind velocity field over 13 years 1979–1991 for the southern North Sea; (a) mean wind velocity field. Location of station EUR is marked with $\circ$; (b) first four EOF modes (from left to right); (c) 8 days (storm No 9, Table 3) extracted from 13 year long 6-hourly time series of first four principal components (first at the top to fourth at the bottom).

In order to verify the results of EOF analysis the comparison of original wind field data $\mathbf{W}_t$ with the wind field reproduced, according to equation (3), by the first four and eight EOF modes and principal components was performed. Examples of this comparison are shown on Figs. 5 (a) and (b). Fig. 5 (a) shows the comparison of time histories of the wind velocity at a selected grid point while Fig. 5 (b) shows the comparison of spatial distribution of the wind velocity field at a selected time instant. It can be seen that reasonably good agreement exists when the first four EOF modes are utilised and becomes very good when the first eight EOF modes are taken into account. The remarkably good agreement reflects the fact that the first eight EOF modes contain 95% of the total variance of analysed wind velocity time series, i.e., total variance calculated using reproduced wind will be equal to 95% of total variance

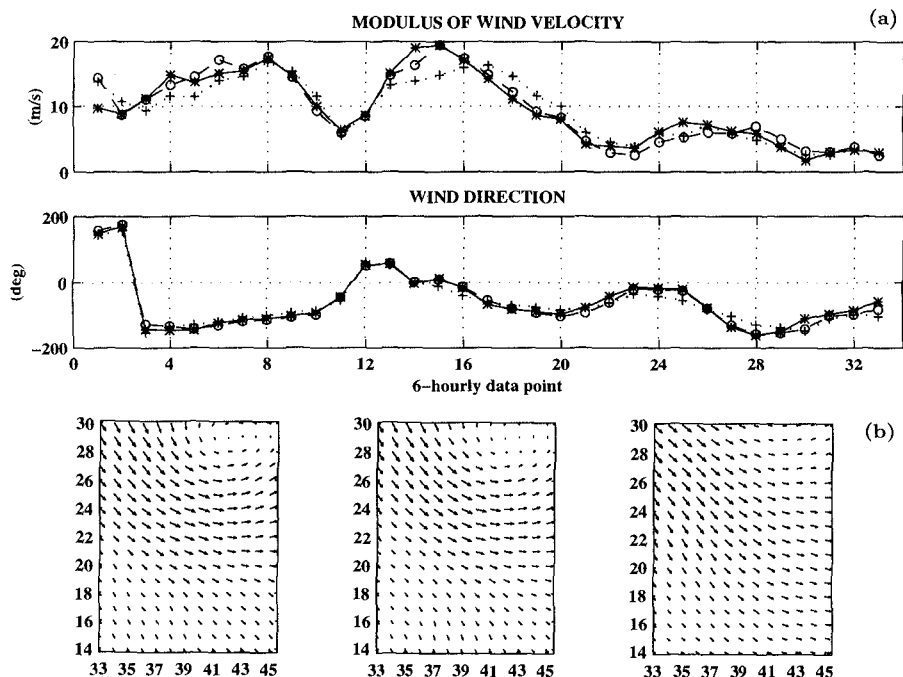


Fig. 5. Results of EOF analysis of wind data over the southern North Sea; (a) comparison of original DNMI wind velocity time series (*) with those recalculated via EOF method using first four (+) and first eight (o) principal components. Time series extending over 8 days (storm No 9, Table 3) are given in grid point (43, 24); (b) spatial distribution of DNMI wind velocity field at 12h on 12 Dec, 1990 (peak of the storm, data point 15); original distribution (to left), and recalculated using first eight principal components (in the middle) and first four (to right).

of the original wind field. This also indicates the scale of reduction in the volume of data that should be handled in further SI procedures. Namely, it shows that using 8 of the 221 EOF modes and 8 of the 221 principal components, one is able to reproduce the information contained in the original data base to a reasonably high level. In other words, using only 3.6% of the whole data set we are still able to retrieve most of the information contained within it.

3. System identification

3.1 Background

The metocean system and basic input-output configuration may be symbolically depicted as in Fig. 6. The output data $X(t)$ across the locations $\mathcal{L}$ are the offshore sea state parameters. The principal components of wind field and atmospheric pressures compose the input data vectors Q_t of N_Q dimension. There is also some unmeasurable random disturbance $\mathcal{N}_t$ that influences the output. We shall include that disturbance as an additive filtered white-noise $\mathcal{E}_t$.

We assume in this study that the metocean system can be described by a linear time-invariant model which is specified by the sequence of impulse response series

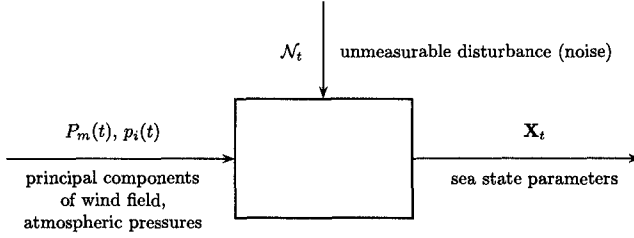


Fig. 6. Basic input-output configuration of metocean system being modelled.

$g_q(k)$, $q = 1, \dots, N_Q$ and the weighting function $h(k)$ of random additive disturbance, $k = 0, 1, \dots, \infty$, and, possibly, the probability density function of the white-noise $\mathcal{E}_t$.

It is worth noting, that despite the assumption of linear model we are still able to incorporate in a system the nonlinearities that have the character of a static nonlinearity at the input side, while dynamics itself is linear. In case the nonlinearity is known, say as function F , the input can be transformed as $Y(t) = F(X(t))$ and the system can be treated as linear. We have such a situation in this study where the moduli of the wind field principal components are taken as stimuli rather than the principal components themselves.

A complete model is given by the following relationship (see e.g. Ljung 1987)

$$X(t) = \mathbf{G}(f)\mathbf{Q}(t) + \mathcal{N}(t) \quad (6)$$

in which f is the forward shift operator, $\mathbf{G}$ is the transfer function of the system and $\mathbf{G}(f)\mathbf{Q}(t)$ is short for

$$\sum_{q=1}^{N_Q} G_q(f)Q_q(t) = \sum_{k=0}^{\infty} \sum_{q=1}^{N_Q} g_q(k)Q_q(t-k) \quad (7)$$

and for any $q = 1, 2, \dots, N_Q$

$$G_q(f) = \sum_{k=0}^{\infty} g_q(k)f^{-k}; \quad f^{-1}Q_q(t) = Q_q(t-1) \quad (8)$$

As mentioned above, we assume that the disturbance $\mathcal{N}$ can be described as filtered white-noise, so

$$\mathcal{N}(t) = H(f)\mathcal{E}(t) \quad (9)$$

where

$$H(f) = 1 + \sum_{k=1}^{\infty} h(k)f^{-k} \quad (10)$$

Within SI we work with the structures that permit the specification of $\mathbf{G}$ and H in terms of a finite number of numerical values. As it is common, we assume that $\mathcal{E}(t)$ is Gaussian, in which case the PDF is specified by the first and second moments. Thus, a particular model (6) is entirely determined in terms of a number of numerical

coefficients which are included as parameters to be determined. The purpose of SI is to determine the values of those parameters. If we denote the parameters in question by the vector θ , and if we take into account equation (9), the basic description for the modelled system becomes

$$X(t) = \mathbf{G}(f, \theta)\mathbf{Q}(t) + H(f, \theta)\mathcal{E}(t) \quad (11)$$

$p_{\mathcal{E}}(\cdot, \theta)$, the PDF of $\mathcal{E}(t)$; $\mathcal{E}(t)$ white-noise

which is a set of models, each of them associated with a parameter value θ .

One of a commonly used way of parameterising G_q and H is to represent them as rational functions of f^{-1} and specify the numerator and denominator coefficient in some way (see e.g. Ljung 1987). Such model structures, which are known as *black-box models*, were utilised in this study. A few different model structures were tested. However, in the limited frame of the present paper we are not able to present the modelling results for all of them. The examples demonstrating the results of SI with the simplest ARX model structures for the significant wave height and surge will be presented. Those models were estimated using *least-squares methods*.

ARX model structure

If in (11) we assume

$$G_q(f, \theta) = \frac{B_q(f)}{A(f)} \quad \text{for } q = 1, \dots, N_Q, \quad H(f, \theta) = \frac{1}{A(f)} \quad (12)$$

where

$$A(f) = 1 + a_1 f^{-1} + \dots + a_{N_A} f^{-N_A} \quad (13)$$

and for $q = 1, 2, \dots, N_Q$

$$B_q(f) = b_0^q + b_1^q f^{-1} + b_2^q f^{-2} + \dots + b_{N_{B_q}}^q f^{-N_{B_q}} \quad (14)$$

we obtain one of the simplest model structures, i.e., the autoregressive with extra input model (ARX). If (11) is rewritten as $A(f)X(t) = \mathbf{B}(f)\mathbf{Q}(t) + \mathcal{E}(t)$, $A(f)X(t)$ is the autoregressive part while $\mathbf{B}(f)\mathbf{Q}(t)$ is the extra input of the ARX model.

The vectorial parameter θ to be determined is in this case

$$\theta = [a_1, a_2, \dots, a_{N_A}, b_0^1, b_1^1, \dots, b_{N_{B_1}}^1, b_0^2, b_1^2, \dots, b_{N_{B_2}}^2, \dots, b_0^{N_Q}, b_1^{N_Q}, \dots, b_{N_{B_{N_Q}}}^{N_Q}] \quad (15)$$

where $N_A, N_{B_1}, N_{B_2}, \dots, N_{B_{N_Q}}$ are the orders of the multi-input ARX model.

By substituting (12) into (11) the following input-output relationship is obtained

$$X(t) + a_1 X(t-1) + \dots + a_{N_A} X(t-N_A) = b_0^1 Q_1(t) + b_1^1 Q_1(t-1) + \dots + b_{N_{B_1}}^1 Q_1(t-N_{B_1}) + \dots + b_0^{N_Q} Q_{N_Q}(t) + b_1^{N_Q} Q_{N_Q}(t-1) + \dots + b_{N_{B_{N_Q}}}^{N_Q} Q_{N_Q}(t-N_{B_{N_Q}}) + \mathcal{E}(t) \quad (16)$$

which is the *linear difference equation*. The ARX model represented by (16) is sometimes called an *equation error model* because of the way in which the white-noise term $\mathcal{E}(t)$ enters the difference equation (16).

3.2 Application in modelling of significant wave height and surge

As mentioned in the introduction, we use the RIKZ measurements as the output data for the purpose of SI. Those 3-hourly measurements covering the 13 year period 1979–1991 were recorded at five offshore monitoring stations.

In this paper, as examples demonstrating the SI carried out within the study, the modelling of significant wave height $H_S(t)$ and surge $S(t)$ in the location of the station EUR are presented. The position of the EUR station is marked on Fig. 4 (a).

Table 3. Selected extreme storms.

Storm no.	Year	Start (00:00 hrs)	End (24:00 hrs)
1	1981	21 Nov	28 Nov
2	1982	13 Dec	18 Dec
3	1983	15 Jan	22 Jan
4	1983	29 Jan	5 Feb
5	1984	11 Jan	18 Jan
6	1989	11 Feb	18 Feb
7	1990	22 Jan	29 Jan
8	1990	23 Feb	2 Feb
9	1990	9 Dec	16 Dec

Within the NEPTUNE project 41 storms over the Irish Sea and the southern North Sea were selected in order to study in a very extensive way the historical extreme coastal events. 9 of the 20 storms selected for the southern North Sea region overlap the 13 year period of the Dutch measurements. They are listed in Table 3. We use those 9 storms to present the results of SI. Namely, both the synthetic data produced by the estimated models and the measured data are plotted and the standard deviations of differences between measured and modelled values are given for the 9 storms listed in Table 3.

In order to select the structure of the model, i.e., to set up the input-vectors and decide about the combination of model orders, a *cross-validation* procedure was utilised. Namely, the data series covering the 13 year period were split into independent *working* and *validation* data sets. The working data (covering 7 years) were used for the estimation while the validation data (covering 6 years) were used to evaluate an estimated model's properties. That evaluation was done mainly by comparison of the simulated and measured output time series, and computing of the sum of squared prediction errors. The model structure resulting in the smallest sum was selected and then estimated again using the whole 13 year long data set. The synthetic data produced by those re-estimated models, for the 9 storms listed in Table 3, are shown in Figs. 7 and 8.

Below we describe the construction of the input-vectors used in SI procedures applied in the modelling of the significant wave height $H_S(t)$ and surge $S(t)$.

Significant wave height

In the modelling of significant wave height H_S the principal components of the wind velocity field for the 13 year period were taken as the only system stimuli. The units of the wind velocity field principal components P_m are, as in the case of wind velocity itself, m/s (EOF modes e_m are dimensionless). The dimensional analysis applied to the significant wave height H_S and principal components P_m suggests that the squares of the latter should be used as input in SI. However, slightly better results were obtained when $|P_m|$ to the first power were also incorporated.

The analysis of equation (3) together with the examination of the spatial patterns shown by the first and second EOF modes with reference to the geographical features of the southern North Sea led to the conclusion that the first EOF mode should have

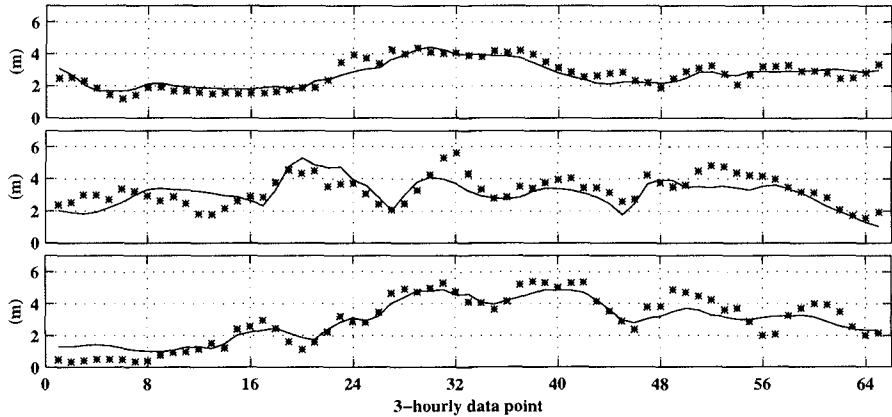


Fig. 7. Comparison between modelled (solid line) and recorded (stars) time series of significant wave height $H_S(t)$ for 3 of the 9 storms listed in Table 3: storms No 1, 5 and 8 (from top to the bottom).

greatest influence when it is rotated by an angle of about $\alpha_1 = -70^\circ$ and also $\beta_1 = 45^\circ$, while for the second EOF mode, the rotation angle of about $\alpha_2 = -80^\circ$ was selected. As a result, the projections of the first principal component P_1 on α_1 and β_1 , and the second principal component P_2 on α_2 were used in the construction of the input-vector rather than their arguments. The projection directions were then subjected to the cross-validation procedure (by varying the projection angles over the ranges around the initially selected values) which confirmed the goodness of the initial guess. In addition, the cross-validation procedure showed it was useful to introduce the second projection also for the second principal component P_2 , namely $\beta_2 = 35^\circ$.

Interestingly, it was found by cross-validation, with the projections of P_1 and P_2 taken into account, that there was no improvement arising from the greater number of components involved in the input-data vector. However, by taking the higher model orders, the modelling was improved, namely, when 7 past input values were taken into account. This suggests that the most important input information in the modelling of the sea response is involved in the first two principal components, despite the fact that they compose only about 75% of the total variance. However, longer history of the dynamic metocean system (21 hours in our case) should be taken into account.

Finally, the following input-vector $\mathbf{Q}^{H_S}$ was selected to model the significant wave height:

$$Q_q^{H_S} = \begin{cases} |P_q| & \text{for } q = 1, 2 \\ |P_{q-2}|^2 & \text{for } q = 3, 4 \\ \Re(P_{q-4} \exp(-i\alpha_{q-4})) & \text{for } q = 5, 6 \\ \Re(P_{q-6} \exp(-i\beta_{q-6})) & \text{for } q = 7, 8 \end{cases} \quad (17)$$

in which $\Re(\cdot)$ denotes the real part of a complex number enclosed in the brackets, with the orders:

$$N_A = 0; \quad N_{B_1} = N_{B_2} = \dots = N_{B_8} = 7 \quad (18)$$

Fig. 7 demonstrates the result of ARX modelling with the structure given by (17) and (18). The standard deviation of the differences between the synthetic and measured significant wave heights are listed in Table 4.

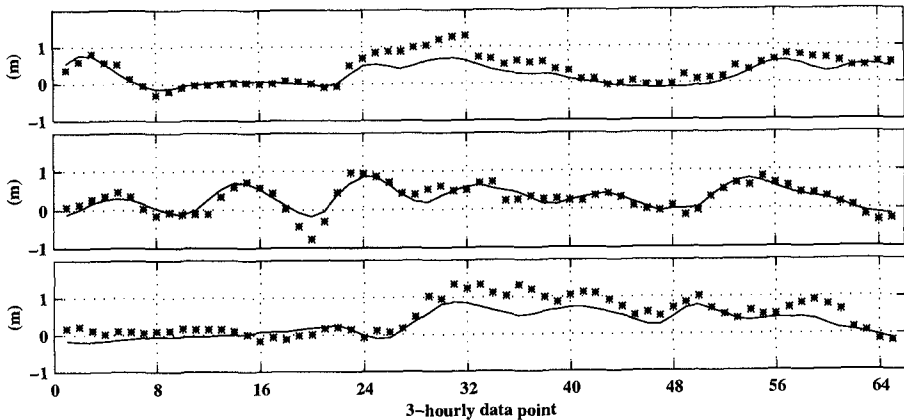


Fig. 8. Comparison between modelled (solid line) and recorded (stars) time series of surge $S(t)$ for 3 of the 9 storms listed in Table 3: storms No 1, 5 and 8 (from top to the bottom).

Surge

In its simplest form the SI can be considered to take the form of regression techniques applied in past years to examine and estimate storm surges (see e.g. Ovadia 1980 and Amin 1982). In this section the results of surge modelling using the ARX structure with the principal components of wind velocity field and local atmospheric pressure taken as the system stimuli are presented. The dimensional analysis applied to the surge S , principal components P_m and local pressure p suggests to use the squares of principal components and the atmospheric pressure to the first power. The input-

Table 4. Standard deviations of differences between modelled and recorded time series.

Calculated for	Variable and model used	
	H_S	S
	ARX (17) & (18)	ARX (19) & (20)
Storm 1	0.37	0.19
Storm 2	0.45	0.13
Storm 3	0.52	0.16
Storm 4	0.51	0.24
Storm 5	0.69	0.16
Storm 6	0.48	0.28
Storm 7	0.55	0.14
Storm 8	0.60	0.22
Storm 9	0.66	0.29
Average of 9 storms	0.54	0.20
13 years: 1979–1991	0.46	0.12

vector $\mathbf{Q}^S$ was selected in the same manner as for significant wave height H_S (with the first two principal component projections) with additional component being the local atmospheric pressure. However, in this case, contrary to H_S case, the cross-validation analysis suggested not to take into account the first powers of the principal component moduli. Finally, the input-data vector $\mathbf{Q}^S$ was constructed as

$$Q_q^{H_S} = \begin{cases} |P_q|^2 & \text{for } q = 1, 2 \\ \Re(P_{q-2} \exp(-i\alpha_{q-2})) & \text{for } q = 3, 4 \\ \Re(P_{q-4} \exp(-i\beta_{q-4})) & \text{for } q = 5, 6 \\ p & \text{for } q = 7 \end{cases} \quad (19)$$

with the orders:

$$N_A = 0; \quad N_{B_1} = N_{B_2} = \dots = N_{B_7} = 7 \quad (20)$$

The local pressure p was computed from the DNMI pressure data base by bi-linear interpolation of the time series taken from the four neighbour DNMI grid points.

The surge time series created by the ARX model with the structure given in (19) and (20) are demonstrated in Fig. 8. The standard deviations of errors are listed in Table 4.

4. Conclusions

Comparison between the modelled and observed time series of both significant wave height and surge level present a sufficiently good agreement to prove effectiveness of the new approach. In the case of significant wave height for 9 storms examined, the standard deviations of differences between the predicted and observed values appeared to be of the same order as those found for the second generation numerical wave forecasting model utilised within the project. Extreme values are found to be underestimated which may be due to possible inconsistency between the input and output data, i.e., the DNMI synthetic wind field and the Dutch field measurements. Further research is being directed to improve the modelling of individual extremes.

This study was carried out as a part of the NEPTUNE project founded by the Commission of the European Communities, Directorate General for Science, Research, and Development under MAST contract No. MAS2-CT94-0081. The Royal Norwegian Meteorological Institute and the National Institute for Coastal and Marine Management (RIKZ) provided data utilised in this study. We are especially grateful to Dr John de Ronde of RIKZ for his permission to access the Dutch offshore station data.

References

- Amin, M. 1982. On analysis and forecasting of surges on the west coast of Great Britain. *Geophys. J. R. astr. Soc.*, **68**, 79–94.
- Graff, J. & Cieřlikiewicz, W. 1996. Determining NW European coastal extremes within the NEPTUNE project. *Proc. 25th Coastal Engng Conf., ICCE'96*, Orlando, Florida.
- Graff, J., Cieřlikiewicz, W., de Haan, L., de Ronde, J., Holt, T., Hurdle, D., Rosenthal, W., & Tawn, J. 1995. NEPTUNE—An integrated approach to determining NW European coastal extremes. *Proc. Second MAST Days and EUROMAR Market*, Sorrento, **1**, 685–700.
- Ljung, L. 1987. *System identification: theory for the user*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Ng, B. 1993. Tidal current predictions using rotary empirical orthogonal functions. *J. Atmos. Oceanic Technol.*, **10**, No. 6, 868–879.
- Ovadia, D. C. 1980. A regression model for the statistical prediction of extreme sea levels at Liverpool. *Rep. No. 102*, Institute of Oceanographic Sciences (unpublished manuscript).
- Prandle, D. & Matthews, J. 1990. The dynamics of nearshore surface currents generated by tides, wind and horizontal density gradients. *Contin. Shelf Res.*, **10**, 665–681.